

Fletcher-Turek Model Averaged Profile Likelihood Confidence Intervals

Paul Kabaila

Department of Mathematics and Statistics, La Trobe University

A.H. Welsh

Mathematical Sciences Institute, The Australian National University

Waruni Abeysekera

Department of Mathematics and Statistics, La Trobe University

April 29, 2014

Running headline: Model Averaged Confidence Intervals

Abstract

We evaluate the model averaged profile likelihood confidence intervals proposed by Fletcher and Turek (2011) in a simple situation in which there are two linear regression models over which we average. We obtain exact expressions for the coverage and the scaled expected length of the intervals and use these to compute these quantities in particular situations. We show that the Fletcher-Turek confidence intervals can have coverage well below the nominal coverage and expected length greater than that of the standard confidence interval with coverage equal to the same minimum coverage. In these situations, the Fletcher-Turek confidence intervals are unfortunately not better than the standard confidence interval used after model selection but ignoring the model selection process.

Keywords: Akaike Information Criterion (AIC); confidence interval; coverage probability; expected length; model selection; nominal coverage, regression models.

1 Introduction

It is common practice in applied statistics to carry out data-based model selection by, for example, using preliminary hypothesis tests or minimizing a criterion such as the Akaike Information Criterion (AIC) and then to use the selected model to construct confidence intervals as if it had been given to us *a priori* as the true model. This procedure can lead to confidence intervals with minimum coverage probabilities far below the nominal coverage probability; see Kabaila (2009) for a review of the literature on this topic.

In recent years, there has been growing interest in using techniques which involve several models to try to incorporate model uncertainty into the inferences. These techniques, loosely referred to as model-averaging, are used in both the Bayesian and the frequentist literature; see, for example, Buckland et al. (1997), Raftery et al. (1997), Volinsky et al. (1997), Hoeting et al. (1999), Burnham and Anderson (2002) and Claeskens and Hjort (2008). In this paper, we focus on frequentist model-averaging techniques for constructing confidence intervals.

The earliest frequentist approach to constructing model-averaged confidence intervals (see Buckland et al, 1997 and Burnham and Anderson, 2002) was to centre the interval on a model-averaged estimator and determine the width of the interval by an estimate of the standard deviation of this estimator. The distribution theory on which these intervals are based is not (even approximately) correct (Claeskens and Hjort, 2008, p.207) but simulation studies report that these intervals work well in terms of coverage probability in particular cases (Lukacs et al., 2010; Fletcher and Dillingham, 2011). A different approach was proposed by Hjort and Claeskens (2003) but this turns out to be essentially the same as the standard confidence interval based on fitting a full model (Kabaila and Leeb, 2006; Wang and Zou,

2013). More recently, Fletcher and Turek (2011) and Turek and Fletcher (2012) have proposed averaging confidence interval construction procedures from each of the possible models. Fletcher and Turek (2011) averaged the profile likelihood confidence interval procedure and Turek and Fletcher (2012) averaged the tail areas of the distributions of the estimators from each of the possible models.

Given the practical importance of the problem, it is not surprising that considerable hope has been invested in model averaging as a simple, general method for making valid inferences under model uncertainty. In this context, it is important to develop a theoretical understanding of the properties of model averaging procedures so that we can put their increasing use on a firm basis. A good starting point is to explore the properties of procedures in meaningful, tractable scenarios which allow us to evaluate whether they work as expected, to compare different proposals and perhaps to modify and improve current proposals. We make a start on this by studying the theoretical properties of the Fletcher and Turek (2011) model averaged profile likelihood confidence interval procedure in a simple scenario that is both meaningful and tractable.

We obtain a $1 - \alpha$ level profile likelihood confidence interval for a parameter θ in a model \mathcal{M}_j by computing the signed-root log-likelihood ratio for θ under \mathcal{M}_j and then solving for the lower and upper endpoints of the interval the two equations obtained by equating the normal cumulative distribution function evaluated at the signed-root log likelihood ratio to $1 - \alpha/2$ and $\alpha/2$, respectively. When we have models $\{\mathcal{M}_1, \dots, \mathcal{M}_R\}$ for a fixed, finite R , the Fletcher and Turek (2011) model averaged profile likelihood confidence interval (MPI) for θ , with nominal coverage $1 - \alpha$, is obtained by solving for the endpoints a weighted average of the profile likelihood confidence interval endpoint equations for each model. There are various ways to choose the weights; Fletcher and Turek focus on weights derived by exponentiating the Akaike Information Criterion (AIC) for each model.

The only evaluation of the MPI to date has been by simulation; Fletcher and Turek (2011) showed that the MPI performs well in particular settings. It is natural

to use simulations to evaluate different confidence intervals, but simulation methods have weaknesses for evaluating performance criteria. First, simulations cover only a limited set of particular settings (particularly, values of the unknown nuisance parameters) and the conclusions apply only to these settings. They may therefore not consider settings where the coverage is low or the expected length is large. We can improve the situation by evaluating minimum coverage probabilities and maximum expected lengths to characterise performance over unknown nuisance parameters. Secondly, the variability in simulation results complicates finding bounds on coverage or expected length, particularly when there are a large number of parameters to vary in the underlying distribution. We therefore use exact calculations to evaluate the properties of the confidence intervals both in particular settings and uniformly over unknown nuisance parameters.

For simplicity, we consider a scenario with only two possible models, a linear regression model with independent and identically distributed normal errors (\mathcal{M}_2) and the same model with a linear constraint on the regression parameters (\mathcal{M}_1). We evaluate the properties of the MPI, with nominal coverage $1 - \alpha$, for a parameter of interest θ that is common to both models. This scenario is simple but, nonetheless, includes practically important problems. For example, in the comparison of two treatments for a given value of the single covariate in a one-way analysis of covariance, the parameter of interest θ is the treatment effect for a given value of the covariate and the two models \mathcal{M}_2 and \mathcal{M}_1 are distinguished by whether τ , the difference in the coefficients of the covariate, is unconstrained or constrained to equal zero (so the fitted models have parallel mean functions). In general, θ and τ can be any linearly independent linear functions of the regression parameter and we obtain general results for any given model matrix, so allowing any possible set of nuisance regression parameters. We focus on two properties, the coverage and the scaled expected length, where the scaling is with respect to the length of the standard confidence interval at the minimum coverage level. We derive computationally convenient, exact expressions for the coverage probability and the scaled expected

length of the MPI for θ , so that we do not need to resort to simulations.

Our results show that there are situations in which the MPI would be expected to work well but has poor coverage, much lower than the nominal coverage, and expected length greater than that of the standard confidence interval with coverage equal to the minimal coverage. In these situations, the MPI performs worse than standard confidence intervals used after model selection but ignoring the model selection process. While disappointing result undermines the hope that the MPI could be generally applicable, it reinforces the the need to develop new procedures and highlights the need for careful analysis of new procedures.

We present our theoretical results in Section 2 and illustrate their application to a real data example from a cloud seeding experiment in which the parameter of interest is the effect of cloud seeding in Section 3. We present the coverage probability and the scaled expected length of the MPI for the parameter of interest and show how to interpret these values. We conclude with a brief discussion in Section 4. Theoretical calculations and the proofs of the Theorems are presented in an Appendix.

2 Theoretical details

In this Section, we describe how to compute the profile likelihood confidence interval for θ and the MPI for θ , and then give exact theoretical expressions for the coverage and the scaled expected length of these intervals. The proofs are left to the Appendix.

The model \mathcal{M}_2 is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where \mathbf{Y} is a random n -vector of responses, \mathbf{X} is a known $n \times p$ model matrix with $p < n$ linearly independent columns, $\boldsymbol{\beta}$ is an unknown p -vector parameter and $\boldsymbol{\varepsilon} \sim \text{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, with σ^2 an unknown positive parameter. Suppose that we are interested in making inference about the parameter $\theta = \mathbf{a}^\top \boldsymbol{\beta}$, where \mathbf{a} is a specified

nonzero p -vector. Suppose also that we define the parameter $\tau = \mathbf{c}^\top \boldsymbol{\beta} - t$, where \mathbf{c} is a specified nonzero p -vector that is linearly independent of \mathbf{a} and t is a specified number. The model \mathcal{M}_1 is \mathcal{M}_2 with $\tau = 0$.

Let $\hat{\boldsymbol{\beta}}$ be the least squares estimator of $\boldsymbol{\beta}$ and $\hat{\sigma}^2 = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-p)$ be the usual unbiased estimator of σ^2 . Set $\hat{\theta} = \mathbf{a}^\top \hat{\boldsymbol{\beta}}$ and $\hat{\tau} = \mathbf{c}^\top \hat{\boldsymbol{\beta}} - t$. Define $v_\theta = \mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a}$ and $v_\tau = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c}$. Then two important quantities are the known correlation $\rho = \mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c} / (v_\theta v_\tau)^{1/2}$ between $\hat{\theta}$ and $\hat{\tau}$ and the unknown parameter $\gamma = \tau / (\sigma v_\tau^{1/2})$.

We adopt the definition of signed-root log-likelihood ratio statistic used by Fletcher and Turek (2011). This is *minus* the usual definition; which definition we adopt makes no essential difference to the results. We show in the Appendix that the signed-root log-likelihood ratio statistic for \mathcal{M}_2 is $r_2\{(\hat{\theta} - \theta)/v_\theta^{1/2}, \hat{\sigma}\} = r_2\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \hat{\sigma}/\sigma\}$, where

$$r_2(\delta, y) = \text{sign}(\delta) \left[n \log \left\{ 1 + \frac{\delta^2}{(n-p)y^2} \right\} \right]^{1/2}, \quad (1)$$

and the signed-root log-likelihood ratio statistic for \mathcal{M}_1 is $r_1\{(\hat{\theta} - \theta)/v_\theta^{1/2}, \hat{\tau}/v_\tau^{1/2}, \hat{\sigma}\} = r_1\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\}$, where

$$r_1(\delta, x, y) = \text{sign}(\delta - \rho x) \left(n \log \left[1 + \frac{(\delta - \rho x)^2}{(1 - \rho^2)\{x^2 + (n-p)y^2\}} \right] \right)^{1/2}. \quad (2)$$

We can derive a profile likelihood confidence interval for θ from the models \mathcal{M}_2 and \mathcal{M}_1 separately or from a weighted average of the profile likelihood confidence endpoint equations for the models \mathcal{M}_2 and \mathcal{M}_1 . Let Φ denote the standard normal cumulative distribution function. Then the profile likelihood confidence interval for θ from model \mathcal{M}_2 , with nominal coverage $1 - \alpha$, is $[\hat{\theta}_{2l}, \hat{\theta}_{2u}]$, where $\hat{\theta}_{2l} < \hat{\theta}$ solves in θ the equation

$$\Phi \left[r_2 \left\{ (\hat{\theta} - \theta)/v_\theta^{1/2}, \hat{\sigma} \right\} \right] = 1 - \alpha/2$$

and $\hat{\theta}_{2u} > \hat{\theta}$ solves in θ the equation

$$\Phi \left[r_2 \left\{ (\hat{\theta} - \theta)/v_\theta^{1/2}, \hat{\sigma} \right\} \right] = \alpha/2.$$

The MPI is obtained by averaging the equations defining the profile likelihood intervals under the models \mathcal{M}_2 and \mathcal{M}_1 . Fletcher and Turek (2011) focus on the Akaike weights which, for the models \mathcal{M}_1 and \mathcal{M}_2 , are w_1 and $1 - w_1$ respectively, where $w_1 = w_1(\hat{\tau}/v_\tau^{1/2}, \hat{\sigma}) = w_1(\hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma)$, with

$$w_1(x, y) = \frac{1}{1 + \left\{1 + \frac{x^2}{(n-p)y^2}\right\}^{n/2} \exp(-1)}. \quad (3)$$

We can consider other weights, including weights obtained by replacing AIC by other model selection criteria. We follow Fletcher and Turek for now and discuss the effect of changing the weights below. For each $x \in R$ and $y > 0$, define

$$h(\delta, x, y) = w_1(x, y) \Phi\{r_1(\delta, x, y)\} + \{1 - w_1(x, y)\} \Phi\{r_2(\delta, y)\} \quad (4)$$

with r_2 , r_1 and w_1 defined by (1)–(3). The MPI, with nominal coverage $1 - \alpha$, is $[\hat{\theta}_l, \hat{\theta}_u]$, where $\hat{\theta}_l < \hat{\theta}$ and $\hat{\theta}_u > \hat{\theta}$ satisfy

$$h\left\{(\hat{\theta} - \hat{\theta}_l)/v_\theta^{1/2}, \hat{\tau}/v_\tau^{1/2}, \hat{\sigma}\right\} = 1 - \alpha/2 \quad \text{and} \quad h\left\{(\hat{\theta} - \hat{\theta}_u)/v_\theta^{1/2}, \hat{\tau}/v_\tau^{1/2}, \hat{\sigma}\right\} = \alpha/2$$

respectively.

We are interested in the coverage and expected length properties of the MPI of Fletcher and Turek (2011). For each $x \in R$ and $y > 0$, define $\delta_u(x, y)$ to be the solution in δ of the equation $h(\delta, x, y) = u$, where h is defined by (4). Theorems 1–3 below on the properties of the MPI are used to construct Figures 1–7 in the next Section; the proofs are given in the Appendix.

Theorem 1 *The coverage probability of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2), with nominal coverage $1 - \alpha$, is*

$$\begin{aligned} P\left(\hat{\theta}_l \leq \theta \leq \hat{\theta}_u\right) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi\left\{\frac{\delta_{1-\alpha/2}(x, y) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}}\right\} \right. \\ &\quad \left. - \Phi\left\{\frac{\delta_{\alpha/2}(x, y) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}}\right\} \right] \phi(x - \gamma) f_{n-p}(y) dx dy, \end{aligned}$$

where ϕ is the probability density function of the standard normal distribution and $f_\nu(y)$ is the probability density function of $(Q/\nu)^{1/2}$, where Q has a χ_ν^2 distribution.

Theorem 1 shows that the coverage of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2) is a function of the nominal coverage $1 - \alpha$, the residual degrees of freedom $n - p = n(1 - p/n)$, the correlation ρ between $\hat{\theta}$ and $\hat{\tau}$, and the unknown parameter $\gamma = \tau / (\sigma v_\tau^{1/2})$. The only unknown quantity is γ . We use the minimum coverage over γ to describe the worst case results without having to specify particular values for γ .

We can relate the coverage probability of the MPI to that of the profile likelihood confidence interval under \mathcal{M}_2 and obtain a very useful upper bound to the minimum coverage probability.

Corollary 1 *The coverage probability of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2), with nominal coverage $1 - \alpha$, converges to the coverage probability of the profile likelihood interval under \mathcal{M}_2 , with nominal coverage $1 - \alpha$, as $\gamma \rightarrow \infty$. That is,*

$$P\left(\hat{\theta}_l \leq \theta \leq \hat{\theta}_u\right) \rightarrow P\left(\hat{\theta}_{2l} \leq \theta \leq \hat{\theta}_{2u}\right) \quad \text{as } \gamma \rightarrow \infty.$$

An immediate consequence is that

$$\inf_{\gamma} P\left(\hat{\theta}_l \leq \theta \leq \hat{\theta}_u\right) \leq P\left(\hat{\theta}_{2l} \leq \theta \leq \hat{\theta}_{2u}\right) = 2 G_{n-p} \left[(n-p)^{1/2} \left\{ \exp\left(\frac{z_{1-\alpha/2}^2}{n}\right) - 1 \right\}^{1/2} \right] - 1,$$

where G_{n-p} denotes the distribution function of the Student t distribution with $n - p$ degrees of freedom and $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$.

Corollary 1 shows that the minimum coverage probability of the MPI cannot be better than the coverage probability of the profile likelihood interval under \mathcal{M}_2 . (Of course, it could be worse.) In effect, if the profile likelihood interval under \mathcal{M}_2 has poor coverage, this will be inherited by the MPI. Perhaps surprisingly, the coverage of the profile likelihood interval can be well below the nominal level $1 - \alpha$. To see this note that for fixed $p/n = r$, the upper bound on the coverage probability is

$$\begin{aligned} & 2 G_{n(1-r)} \left[n^{1/2} (1-r)^{1/2} \left\{ \exp\left(\frac{z_{1-\alpha/2}^2}{n}\right) - 1 \right\}^{1/2} \right] - 1 \\ &= 2 G_{n(1-r)} \left[(1-r)^{1/2} \left\{ z_{1-\alpha/2}^2 + O(n^{-1}) \right\}^{1/2} \right] - 1 \\ &\rightarrow 2\Phi \left\{ (1-r)^{1/2} z_{1-\alpha/2} \right\} - 1, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus the coverage probability of the profile likelihood confidence interval under \mathcal{M}_2 decreases as $p/n = r$ increases and is substantially less than the nominal coverage $1 - \alpha$ unless p/n is small. Corollary 1 shows that the MPI will also have poor coverage properties unless p/n is small.

For the expected length of the MPI, we obtain the following result.

Theorem 2 *The expected length of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2), with nominal level $1 - \alpha$, is*

$$E(\hat{\theta}_u - \hat{\theta}_l) = \sigma v_\theta^{1/2} \int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(x, y) - \delta_{\alpha/2}(x, y)\} \phi(x - \gamma) f_{n-p}(y) dx dy,$$

where ϕ is the probability density function of the standard normal distribution and $f_\nu(y)$ is the probability density function of $(Q/\nu)^{1/2}$, where Q has a χ_ν^2 distribution.

Let c_{\min} denote the minimum coverage probability of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2), with nominal coverage $1 - \alpha$. The scaled expected length of this confidence interval is therefore

$$\begin{aligned} & \frac{E(\hat{\theta}_u - \hat{\theta}_l)}{2 G_{n-p}^{-1}((c_{\min} + 1)/2) E(\hat{\sigma}) v_\theta^{1/2}} \\ &= \frac{E(\hat{\theta}_u - \hat{\theta}_l)}{2 G_{n-p}^{-1}((c_{\min} + 1)/2) \sigma v_\theta^{1/2} \int_0^\infty y f_{n-p}(y) dy} \\ &= \frac{\int_0^\infty \int_{-\infty}^\infty \{\delta_{1-\alpha/2}(x, y) - \delta_{\alpha/2}(x, y)\} \phi(x - \gamma) f_{n-p}(y) dx dy}{2 G_{n-p}^{-1}((c_{\min} + 1)/2) \int_0^\infty y f_{n-p}(y) dy}. \end{aligned}$$

The integral in the denominator has the analytic expression $2^{1/2} \Gamma\{(n-p+1)/2\} / [(n-p)^{1/2} \Gamma\{(n-p)/2\}]$. As with the coverage, the only unknown quantity in this expression is γ , so we study the maximum scaled expected length over γ .

The range of calculations needed to evaluate the coverage probability and the scaled expected length of the MPI are reduced by the following result that shows that, because of symmetry, we need only consider $\gamma \geq 0$ and $\rho \geq 0$.

Theorem 3 *The coverage probability and the scaled expected length of the MPI (averaged over \mathcal{M}_1 and \mathcal{M}_2) are both even functions of γ for fixed ρ and even functions of ρ for fixed γ .*

As we noted above, we can replace AIC by other model selection criteria in the weights. A convenient way to do this is to replace the penalty $2 \times (\# \text{ regression parameters})$ in AIC by $d \times (\# \text{ regression parameters})$, where, for some $0 \leq u \leq 1$,

$$d = n \log \left[1 + \frac{\{G_{n-p}^{-1}(1 - u/2)\}^2}{n - p} \right] \rightarrow z_{1-u/2}^2, \quad \text{as } n \rightarrow \infty.$$

Here u is the significance level of the equivalent test for the significance of an additional parameter; see the Appendix for more details. In this case, the $\exp(-1)$ term in the Akaike weights is replaced by $\exp(-d/2)$. Using the asymptotic approximation to d , we find that AIC corresponds to $u = 0.157$; more extreme examples (lower significance level) are the usual $u = 0.05$ level which gives $\exp(-3.84/2) \approx \exp(-1.92)$ and the Bayesian Information Criterion (BIC) $u = 2[1 - \Phi\{(\log n)^{1/2}\}]$ which gives $\exp\{-\log(n)/2\} = n^{-1/2}$; less extreme examples (higher significance level) such as $u = 0.5$ which gives $\exp(-0.45/2) \approx \exp(-0.227)$ can also be considered.

We explored the effect of changing d , hoping in particular that values of $d < 2$ might improve the performance of the MPI when p/n is not small, but this is not the case and changing d has very little effect. Theoretical support for this conclusion is provided by noting that Corollary 1 holds for any fixed value of $d \geq 0$ so, irrespective of the fixed value of $d \geq 0$, the minimum coverage probability of the MPI, with nominal coverage $1 - \alpha$, cannot exceed the coverage probability of the profile likelihood confidence interval, with nominal coverage $1 - \alpha$, using \mathcal{M}_2 . Our conclusion is that changing d does not change the “ p/n not small” problem. In the boundary case, $d = 0$, we have no penalty on the number of regression parameters so we might expect MPI to always use the model \mathcal{M}_2 . However, the weight reduces to

$$w_1 = \frac{1}{1 + \left(1 + \frac{\hat{\tau}^2}{(n-p)\hat{\sigma}^2 v_\tau}\right)^{n/2}} \quad (5)$$

in this case and we still average over the two models. Similarly, for each fixed $d \geq 0$, we do not recover the profile likelihood confidence interval as $\gamma \rightarrow 0$ or even $\hat{\gamma} \rightarrow 0$, but continue to average over the two models.

3 Cloud seeding example

In this Section, we illustrate how we can use our results on the properties of the MPI in the context of a real data example from a cloud seeding experiment. The data are presented and analysed by Biondini, Simpson and Woodley (1997), Miller (2002, Section 3.12) and Kabaila (2005). Following Kabaila (2005), we compare the effect of seeding (TRT=1) against the random control (TRT=2) treatment in the moving echo motion category (CAT=1) subgroup of the data. The response variable is the floating target rainfall volume ($m^3 \times 10^7$) and the sample size is $n = 33$. In addition to the treatment indicator, there are five other predictor variables: coverage (percent) which measures the cloud cover in the target area; seedability (km); prewetness ($m^3 \times 10^7$) which measures the rainfall in the target area in the hour before treatment; earliness (hrs) which measures the number of hours in the morning in which there were clouds in the target area; and the average speed of echo motion (knots). The models considered by Miller (2002, Section 3.12) and Kabaila (2005) included the intercept, treatment indicator, the main effects, squared effects and the interactions between the five predictor variables so that p , the dimension of the regression parameter vector, is 22. All these additional variables can be included in the model or not; variable selection has been carried out by Miller (2002, Section 3.12) and Kabaila (2005) for many variables in this study. For illustration, we consider model averaging over the full model ($p = 22$) and the submodel excluding the squared seedability term s^2 whose coefficient we denote by τ . The goal is to construct a 95% confidence interval for θ , the expected response when cloud seeding is used minus the expected response under random control when all the other explanatory variables are the same.

We can construct profile-likelihood and MPI (over \mathcal{M}_2 and \mathcal{M}_1) confidence intervals for θ . The standard 0.95 Student t confidence interval for θ under model \mathcal{M}_2 is $[-0.327, 3.421]$, the profile-likelihood confidence interval for θ under model \mathcal{M}_2 , with nominal coverage 0.95, is $[0.554, 2.539]$ and the MPI for θ , with nominal coverage 0.95, is $[0.618, 2.572]$. For comparison, the standard confidence interval for

θ , with nominal coverage 0.95, after selection between models \mathcal{M}_1 and \mathcal{M}_2 using AIC but ignoring the model selection process is $[0.474, 2.650]$. Model averaged profile likelihood confidence intervals for θ are held to be better than the confidence interval that ignores the model selection process, because they should better reflect the uncertainty in choosing between the two models. For the MPI, we plot the exact coverage and the scaled expected length in Figures 1 and 2, respectively. We find that the coverage probability of the MPI is close to 0.7315 for all γ rather than the nominal 0.95 and the scaled expected length is close to one for all γ . Therefore, the MPI is actually similar to the standard 0.7315 confidence interval for θ . This is not quite the good performance hoped for under model averaging.

[Figures 1 and 2 near here]

The important quantities for the MPI based on models \mathcal{M}_2 and \mathcal{M}_1 are p/n and ρ , the correlation between the least squares estimators of θ and τ . For the cloud seeding example, $p/n = 2/3$ which is not small and the correlation between $\hat{\theta}$ and $\hat{\tau}$ (which depends on \mathbf{X} and the choice of θ and τ so is known) is $\rho = 0.2472$ which is small and positive. The minimum coverage against $|\rho|$ for fixed $p/n = 2/3$ is plotted in Figure 3.

[Figure 3 near here]

The coverage properties of the MPI as a function of p/n are inherited from those of the profile likelihood confidence interval. We showed in Section 2 that the minimum coverage probability of the MPI, with nominal coverage $1 - \alpha$, cannot be larger than the coverage probability of the profile likelihood confidence interval under \mathcal{M}_2 , with the same nominal coverage. An asymptotic expansion of the latter coverage probability showed that it will be substantially below $1 - \alpha$, unless p/n is small (obviously, $0 < p/n < 1$). This is confirmed by plotting the coverage of the profile likelihood confidence interval under \mathcal{M}_2 against p/n in Figure 4. The coverage decreases strongly as either $|\rho|$ or p/n increase; in the cloud seeding example, the poor coverage is driven by p/n not being small.

[Figure 4 near here]

It is interesting to compare the MPI interval with the naive confidence interval constructed after selecting between models \mathcal{M}_1 and \mathcal{M}_2 the model with smaller AIC and treating the selected model as if it had been given to us *a priori* as the true model. The coverage probability of this interval as a function of γ is shown in Figure 5 (Kabaila and Giri, 2009a, b). Comparing this with Figure 1, we see that the coverage probability for this naive post-model-selection interval is uniformly far better than that of the MPI.

[Figure 5 near here]

For a second example, suppose that we change τ from the coefficient of the squared seedability to the seedability-earliness interaction. In this case, n and p are unchanged but now $\rho = -0.4530$. The MPI for θ , with nominal coverage 0.95, is $[0.689, 2.540]$, which is quite similar to the previous case. We plot the exact coverage and the scaled expected length for the MPI in Figures 6 and 7, respectively. The coverage probability of the MPI is close to 0.728 for all γ rather than the nominal 0.95 and the scaled expected length is close to one for all γ , although the curves are different from those obtained in Figures 1 and 2. We conclude that the MPI has similar coverage and expected length properties to the standard 0.728 confidence interval for θ . The naive confidence interval constructed after selecting between models \mathcal{M}_1 and \mathcal{M}_2 the model with smaller AIC and treating the selected model as if it had been given to us *a priori* as the true model has similar coverage to that shown in Figure 5. Once again, we see that the coverage probability for the naive post-model-selection interval is uniformly far better than that of the MPI.

[Figures 6 and 7 near here]

4 Conclusion

We have examined the exact coverage and scaled expected length of the MPI for a parameter θ , with nominal coverage $1 - \alpha$, in a particular simple situation in which there are two linear regression models (differing in only a single parameter τ) to average over. We showed that both the coverage and the scaled expected length depend on n , $n - p$, the correlation ρ between the least squares estimators $\hat{\theta}$ and $\hat{\tau}$, and the unknown true value $\gamma = \tau/(\sigma v_\tau^{1/2})$. As γ is unknown, it is useful to consider the minimum coverage and the maximum scaled expected length over γ . The results show that the MPI can perform poorly when p/n is not small or when $|\rho|$ is large, and should not be used in these situations. In fact, in these situations, the MPI performs no better than than post model selection confidence intervals which ignore the selection process.

The MPI is obtained by trying to average profile likelihood confidence intervals and we have shown that the performance of the MPI is limited by the performance of the underlying profile likelihood confidence intervals. In particular, the MPI inherits poor performance when p/n is not small from the fact that profile likelihood confidence intervals perform poorly when p/n is not small. Averaging other types of confidence intervals which do not have this problem may lead to better results, at least when p/n is not small.

References

- Biondini, R., Simpson, J. and Woodley, W. (1977). Empirical predictors for natural and seeded rainfall in the Florida area cumulus experiment (FACE), 1970-1975. *Journal of Applied Meteorology* **16**, 585–594.
- Buckland, S.T., Burnham, K.P. and Augustin, N.H. (1997). Model selection: an integral part of inference. *Biometrics* **53**, 603–618.
- Burnham, K.P. and Anderson, D.R. (2002). *Model selection and multimodel in-*

- ference, a practical information-theoretic approach, 2nd edition. Springer: New York.
- Claeskens, G. and Hjort, N.L. (2008). *Model selection and model averaging*. Cambridge University Press.
- Fletcher, D. and Dillingham, P.W. (2011). Model-averaged confidence intervals for factorial experiments. *Comput. Statist. Data Anal.* **55**, 3041–3048.
- Fletcher, D. and Turek, D. (2011). Model-averaged profile likelihood intervals. *J. Agric. Biol. Environ. Stat.* **17**, 38–51.
- Hoeting, J., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999). Bayesian model-averaging: A tutorial (with discussion). *Statist. Sci.* **14**, 382–417.
- Hjort, N.L. and Claeskens, G. (2003). Frequentist model average estimators. *J. Amer. Statist. Assoc.* **98**, 879–899.
- Kabaila, P. (2005). On the coverage probability of confidence intervals in regression after variable selection. *Aust. N. Z. J. Stat.* **47**, 549–562.
- Kabaila, P. (2009). The coverage properties of confidence regions after model selection. *International Statistical Review* **77**, 405–414.
- Kabaila, P. and Leeb, H. (2006). On the large-sample minimal coverage probability of confidence intervals after model selection. *J. Amer. Statist. Assoc.* **101**, 619–629.
- Kabaila, P. and Giri, K. (2009a). Upper bounds on the minimum coverage probability of confidence intervals in regression after model selection. *Aust. N. Z. J. Stat.* **51**, 271–287.
- Kabaila, P. and Giri, K. (2009b). Confidence intervals in regression utilizing uncertain prior information. *J. Statist. Plann. Inference* **139**, 3419–3429.
- Lukacs, P.M., Burnham, K.P. and Anderson, D.R. (2010). Model selection bias and Freedman’s paradox. *Ann. Inst. Statist. Math.* **62**, 117–125.
- Miller, A. (2002). *Subset selection in regression*, 2nd edition. Chapman & Hall/CRC.

Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997). Bayesian model-averaging for linear regression models. *J. Amer. Statist. Assoc.* **92**, 179–191.

Turek, D. and Fletcher, D. (2012). Model-averaged Wald confidence intervals. *Comput. Statist. Data Anal.* **56**, 2809–2815.

Volinsky, C.T., Madigan, D., Raftery, A.E. and Kronmal, R.A. (1997). Bayesian model-averaging in proportional hazard models: Assessing the risk of a stroke. *J. R. Statist. Soc. Ser. C Appl. Stat.* **46**, 433–448.

Wang, H. and Zou, S.Z.F. (2013). Interval estimation by frequentist model averaging. *Commun. Statist. Theory Methods* to appear.

Paul Kabaila, Department of Mathematics and Statistics, La Trobe University, Victoria 3086, Australia.

E-mail: P.Kabaila@latrobe.edu.au

Appendix

The models

It simplifies the presentation if we reparametrise the models \mathcal{M}_1 and \mathcal{M}_2 to be explicit functions of the parameters θ and τ . Let \mathbf{M} be the $p \times p$ matrix with first two rows given by $\mathbf{a}^\top(\mathbf{X}^\top \mathbf{X})^{-1/2}$ and $\mathbf{c}^\top(\mathbf{X}^\top \mathbf{X})^{-1/2}$, respectively, and the remaining $p - 2$ rows given by orthonormal p -vectors that are orthogonal to both \mathbf{a} and \mathbf{c} . The incorporation of $(\mathbf{X}^\top \mathbf{X})^{-1/2}$ into the first two rows of \mathbf{M} may seem unnecessary but in fact, as we will see below, it produces a useful standardisation. The model \mathcal{M}_2 can be written as

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\boldsymbol{\eta} + \boldsymbol{\varepsilon},$$

where $\tilde{\mathbf{Y}} = \mathbf{Y} - t\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1/2}\mathbf{M}^{-1}\mathbf{e}_2$, $\tilde{\mathbf{X}} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1/2}\mathbf{M}^{-1}$ and $\boldsymbol{\eta} = \mathbf{M}(\mathbf{X}^\top \mathbf{X})^{1/2}\boldsymbol{\beta} - t\mathbf{e}_2$, with \mathbf{e}_2 a p -vector with the second component equal to one and all other com-

ponents equal to zero. Write $\boldsymbol{\eta} = (\theta, \tau, \boldsymbol{\psi}^\top)^\top$, where $\boldsymbol{\psi}$ is the $(p-2)$ -vector of the remaining regression parameters. The model \mathcal{M}_1 is \mathcal{M}_2 with $\tau = 0$.

The likelihood for the models

We can write down the log-likelihood for the reparametrised model directly and then re-express it in terms of the maximum likelihood estimators of the parameters, which are a minimal sufficient statistic for \mathcal{M}_2 and \mathcal{M}_1 . It is simpler to first reduce the data and work from the sampling distribution of this minimal sufficient statistic. The maximum likelihood estimator of $\boldsymbol{\eta}$ is given by

$$\hat{\boldsymbol{\eta}} = (\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}})^{-1} \widetilde{\mathbf{X}}^\top \widetilde{\mathbf{Y}} = \mathbf{M}(\mathbf{X}^\top \mathbf{X})^{-1/2} \mathbf{X}^\top \mathbf{Y} - t\mathbf{e}_2$$

and the maximum likelihood estimator of σ^2 is $(n-p)\hat{\sigma}^2/n$, where

$$\hat{\sigma}^2 = (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\eta}})^\top (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{X}}\hat{\boldsymbol{\eta}})/(n-p) = (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})/(n-p).$$

We have

$$\hat{\boldsymbol{\eta}} \sim N \left(\begin{bmatrix} \theta \\ \tau \\ \boldsymbol{\psi} \end{bmatrix}, \sigma^2 \begin{bmatrix} v_\theta & \rho(v_\theta v_\tau)^{1/2} & \mathbf{0}^\top \\ \rho(v_\theta v_\tau)^{1/2} & v_\tau & \mathbf{0}^\top \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{p-2} \end{bmatrix} \right),$$

where $v_\theta = \mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{a}$, $v_\tau = \mathbf{c}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c}$ and $\rho = \mathbf{a}^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{c} / (v_\theta v_\tau)^{1/2}$ are known quantities, and, independently,

$$(n-p)\hat{\sigma}^2/\sigma^2 \sim \chi_{n-p}^2.$$

The advantage of incorporating $(\mathbf{X}^\top \mathbf{X})^{-1/2}$ into the first two rows of \mathbf{M} is that the sampling distribution of $\hat{\boldsymbol{\psi}}$ has a very simple covariance structure with unknown parameter σ^2 . We can write down the log-likelihood for \mathcal{M}_2 which (discarding terms which do not depend on the unknown parameters) is

$$\begin{aligned} \ell_2(\theta, \tau, \boldsymbol{\psi}, \sigma^2) &= -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \left[\frac{1}{(1-\rho^2)} \left\{ \frac{(\hat{\theta} - \theta)^2}{v_\theta} + \frac{(\hat{\tau} - \tau)^2}{v_\tau} - 2\rho \frac{(\hat{\theta} - \theta)(\hat{\tau} - \tau)}{v_\theta^{1/2} v_\tau^{1/2}} \right\} \right. \\ &\quad \left. + (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})^\top (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + (n-p)\hat{\sigma}^2 \right], \end{aligned}$$

and hence the log-likelihood for \mathcal{M}_1 is $\ell_1(\theta, \boldsymbol{\psi}, \sigma^2) = \ell_2(\theta, 0, \boldsymbol{\psi}, \sigma^2)$.

We do not have to specify the particular underlying linear regression model or the specific parameters θ and τ . The results below hold for any full-rank linear regression model and for any linear combinations θ and τ of the regression parameter $\boldsymbol{\beta}$.

The signed-root log-likelihood statistic for \mathcal{M}_2

Setting the derivatives of the log-likelihood $\ell_2(\theta, \tau, \boldsymbol{\psi}, \sigma^2)$ with respect to the unknown parameters to zero and solving the resulting estimating equations shows that the maximum likelihood estimators are $\hat{\theta}$, $\hat{\tau}$, $\hat{\boldsymbol{\psi}}$ and $(n-p)\hat{\sigma}^2/n$, respectively, so the maximum value of the log-likelihood is

$$\ell_2(\hat{\theta}, \hat{\tau}, \hat{\boldsymbol{\psi}}, (n-p)\hat{\sigma}^2/n) = -\frac{n}{2} \log \{(n-p)\hat{\sigma}^2/n\} - \frac{n}{2}.$$

Next, holding θ fixed and setting the derivatives of the log-likelihood $\ell_2(\theta, \tau, \boldsymbol{\psi}, \sigma^2)$ with respect to the remaining unknown parameters to zero, we obtain the maximum profile likelihood estimators $\hat{\tau}(\theta) = \hat{\tau} - \rho(v_\tau/v_\theta)^{1/2}(\hat{\theta} - \theta)$, $\hat{\boldsymbol{\psi}}(\theta) = \hat{\boldsymbol{\psi}}$ and $\hat{\sigma}^2(\theta) = \{(\hat{\theta} - \theta)^2/v_\theta + (n-p)\hat{\sigma}^2\}/n$, so the maximum value of the profile log-likelihood is

$$\ell_2\left\{\theta, \hat{\tau}(\theta), \hat{\boldsymbol{\psi}}(\theta), \hat{\sigma}^2(\theta)\right\} = -\frac{n}{2} \log \left[\left\{ \frac{(\hat{\theta} - \theta)^2}{v_\theta} + (n-p)\hat{\sigma}^2 \right\} / n \right] - \frac{n}{2}$$

It follows that the signed root log-likelihood ratio statistic for \mathcal{M}_2 is $r_2\{(\hat{\theta} - \theta)/v_\theta^{1/2}, \hat{\sigma}\} = r_2\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \hat{\sigma}/\sigma\}$, where r_2 is given by (1).

The signed-root log-likelihood statistic for \mathcal{M}_1

The log-likelihood for model \mathcal{M}_1 is

$$\begin{aligned} \ell_1(\theta, \boldsymbol{\psi}, \sigma^2) = & -\frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \left[\frac{1}{(1-\rho^2)} \left\{ \frac{(\hat{\theta} - \theta)^2}{v_\theta} + \frac{\hat{\tau}^2}{v_\tau} - 2\rho \frac{(\hat{\theta} - \theta)\hat{\tau}}{v_\theta^{1/2} v_\tau^{1/2}} \right\} \right. \\ & \left. + (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi})^T (\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}) + (n-p)\hat{\sigma}^2 \right] \end{aligned}$$

Setting the derivatives of the log-likelihood with respect to the unknown parameters to zero and solving the resulting estimating equations shows that the maximum likelihood estimators are $\widehat{\theta} - \rho(v_\theta/v_\tau)^{1/2}\widehat{\tau}$, $\widehat{\psi}$ and $\{\widehat{\tau}^2/v_\tau + (n-p)\widehat{\sigma}^2\}/n$, respectively, so the maximum value of the log-likelihood is

$$\ell_1\left(\widehat{\theta} - \rho(v_\theta/v_\tau)^{1/2}\widehat{\tau}, \widehat{\psi}, \{\widehat{\tau}^2/v_\tau + (n-p)\widehat{\sigma}^2\}/n\right) = -\frac{n}{2} \log [\{\widehat{\tau}^2/v_\tau + (n-p)\widehat{\sigma}^2\}/n] - \frac{n}{2}.$$

Next, holding θ fixed and setting the derivatives of the log-likelihood $\ell_1(\theta, \psi, \sigma^2)$ with respect to the remaining unknown parameters to zero, we obtain the maximum profile likelihood estimators $\widehat{\psi}_1(\theta) = \widehat{\psi}$ and $\widehat{\sigma}^2(\theta) = \left[\frac{1}{(1-\rho^2)} \{(\widehat{\theta} - \theta)^2/v_\theta + \widehat{\tau}^2/v_\tau - 2\rho(\widehat{\theta} - \theta)\widehat{\tau}/(v_\theta v_\tau)^{1/2}\} + (n-p)\widehat{\sigma}^2 \right]/n$, so the maximum value of the profile log-likelihood is

$$\begin{aligned} \ell_1\{\theta, \widehat{\psi}_1, \widehat{\sigma}_1^2(\theta)\} &= -\frac{n}{2} \log \left(\left[\frac{1}{(1-\rho^2)} \left\{ \frac{(\widehat{\theta} - \theta)^2}{v_\theta} + \frac{\widehat{\tau}^2}{v_\tau} - 2\rho \frac{(\widehat{\theta} - \theta)\widehat{\tau}}{v_\theta^{1/2} v_\tau^{1/2}} \right\} + (n-p)\widehat{\sigma}^2 \right] / n \right) - \frac{n}{2} \\ &= -\frac{n}{2} \log \left(\left[\frac{1}{(1-\rho^2)} \left\{ \frac{(\widehat{\theta} - \theta)^2}{v_\theta} + \rho^2 \frac{\widehat{\tau}^2}{v_\tau} - 2\rho \frac{(\widehat{\theta} - \theta)\widehat{\tau}}{v_\theta^{1/2} v_\tau^{1/2}} \right\} + \frac{\widehat{\tau}^2}{v_\tau} + (n-p)\widehat{\sigma}^2 \right] / n \right) - \frac{n}{2}. \end{aligned}$$

It follows that the signed root log-likelihood ratio statistic for \mathcal{M}_1 is $r_1\{(\widehat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \widehat{\tau}/(\sigma v_\tau^{1/2}), \widehat{\sigma}/\sigma\}$, where r_1 is given by (2).

Akaike weights

For $d = 2$, the Akaike Information Criteria (AIC) for the two models are

$$\text{AIC}_2 = n \log\{(n-p)\widehat{\sigma}^2/n\} + dp$$

and

$$\text{AIC}_1 = n \log[\{(\widehat{\tau}^2/v_\tau) + (n-p)\widehat{\sigma}^2\}/n] + d(p-1),$$

respectively, so the weight is

$$\begin{aligned}
w_1 &= \frac{\exp\left\{-\frac{1}{2}(\text{AIC}_1 - \text{AIC}_{\min})\right\}}{\exp\left\{-\frac{1}{2}(\text{AIC}_1 - \text{AIC}_{\min})\right\} + \exp\left\{-\frac{1}{2}(\text{AIC}_2 - \text{AIC}_{\min})\right\}} \\
&= \frac{1}{1 + \exp\left\{\frac{1}{2}(\text{AIC}_1 - \text{AIC}_2)\right\}} \\
&= \frac{1}{1 + \left\{1 + \frac{\hat{\tau}^2}{(n-p)\hat{\sigma}^2 v_\tau}\right\}^{n/2} \exp(-d/2)}.
\end{aligned}$$

This corresponds to the expression (3).

We can calibrate the choice of d by considering the hypothesis test in which we reject model \mathcal{M}_1 in favour of \mathcal{M}_2 when $\text{AIC}_2 < \text{AIC}_1$. When model \mathcal{M}_1 is true, the probability of rejecting \mathcal{M}_1 (i.e. the level of the test) is

$$2 \left(1 - G_{n-p} \left[(n-p)^{1/2} \left\{ \exp\left(\frac{d}{n}\right) - 1 \right\}^{1/2} \right] \right),$$

where G_{n-p} is the cumulative distribution function of the Student t distribution with $n-p$ degrees of freedom. If we set the level of the test equal to u , we find

$$d = n \log \left[1 + \frac{\{G_{n-p}^{-1}(1-u/2)\}^2}{n-p} \right].$$

Expanding the log function and then letting $n \rightarrow \infty$, we find that

$$d = \{G_{n-p}^{-1}(1-u/2)\}^2 + O(n^{-1}) \rightarrow z_{1-u/2}^2,$$

which can also be expressed in terms of the chi-squared distribution with one degree of freedom.

Proof of Theorem 1

The coverage probability of the MPI confidence interval, with nominal coverage $1 - \alpha$, is

$$P\left(\hat{\theta}_l \leq \theta \leq \hat{\theta}_u\right) = 1 - P\left(\theta < \hat{\theta}_l\right) - P\left(\hat{\theta}_u > \theta\right).$$

Now $h(\delta, x, y)$ is an increasing function of δ for fixed x and y so

$$\begin{aligned}
P(\theta < \hat{\theta}_l) &= P\left\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > (\hat{\theta} - \hat{\theta}_l)/(\sigma v_\theta^{1/2})\right\} \\
&= P\left[h\left\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\right\} > 1 - \alpha/2\right] \\
&= P\left[(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\}\right] \\
&= \int_0^\infty \int_{-\infty}^\infty P[(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\} | \hat{\tau}/(\sigma v_\tau^{1/2}) = x, \hat{\sigma}/\sigma = y] \\
&\quad \times \phi(x - \gamma) f_{n-p}(y) dx dy,
\end{aligned}$$

where $\gamma = \tau/(\sigma v_\tau^{1/2})$. Now the distribution of $(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2})$ conditional on $\hat{\tau}/(\sigma v_\tau^{1/2}) = x$ is $N(\rho(x - \gamma), 1 - \rho^2)$, $\hat{\tau}/(\sigma v_\tau^{1/2}) \sim N(\gamma, 1)$ and $\hat{\theta}$ and $\hat{\tau}$ are independent of $\hat{\sigma}$, so

$$\begin{aligned}
&P[(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}\{\hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\} | \hat{\tau}/(\sigma v_\tau^{1/2}) = x, \hat{\sigma}/\sigma = y] \\
&= P\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > \delta_{1-\alpha/2}(x, y) | \hat{\tau}/(\sigma v_\tau^{1/2}) = x\} \\
&= 1 - P\{(\hat{\theta} - \theta)/\sigma v_\theta^{1/2} \leq \delta_{1-\alpha/2}(x, y) | \hat{\tau}/(\sigma v_\tau^{1/2}) = x\} \\
&= 1 - \Phi\left\{\frac{\delta_{1-\alpha/2}(x, y) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}}\right\}
\end{aligned}$$

and hence

$$P(\theta < \hat{\theta}_l) = \int_0^\infty \int_{-\infty}^\infty \left[1 - \Phi\left\{\frac{\delta_{1-\alpha/2}(x, y) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}}\right\}\right] \phi(x - \gamma) f_{n-p}(y) dx dy.$$

Similarly,

$$\begin{aligned}
1 - P(\hat{\theta}_u > \theta) &= P(\theta < \hat{\theta}_u) \\
&= P\left[h\left\{(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}), \hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\right\} > \alpha/2\right] \\
&= P\left[(\hat{\theta} - \theta)/(\sigma v_\theta^{1/2}) > \delta_{\alpha/2}\{\hat{\tau}/(\sigma v_\tau^{1/2}), \hat{\sigma}/\sigma\}\right] \\
&= \int_0^\infty \int_{-\infty}^\infty \left[1 - \Phi\left\{\frac{\delta_{\alpha/2}(x, y) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}}\right\}\right] \phi(x - \gamma) f_{n-p}(y) dx dy.
\end{aligned}$$

Proof of Corollary 1

From the proof of Theorem 1, we can write

$$P(\theta < \hat{\theta}_l) = 1 - P\{h(G, H, W) \leq 1 - \alpha/2\},$$

where h is defined in (4), $G = (\hat{\theta} - \theta)/\sigma v_\theta^{1/2} \sim N(0, 1)$, $H = \hat{\tau}/\sigma v_\tau^{1/2} \sim N(\gamma, 1)$, $(n - p)W^2 = (n - p)\hat{\sigma}/\sigma \sim \chi_{n-p}^2$ and (G, H) and W are independent. From (3), $w_1(H, W)$ converges in probability to 0, as $\gamma \rightarrow \infty$. Since $0 < \Phi(x) < 1$ for all $x \in \mathbb{R}$, this implies by (4) that $h(G, H, W)$ converges in probability to $\Phi\{r_2(G, W)\}$, as $\gamma \rightarrow \infty$. Thus, $h(G, H, W)$ converges in distribution to $\Phi\{r_2(G, W)\}$, as $\gamma \rightarrow \infty$. The cumulative distribution function of $\Phi\{r_2(G, W)\}$, evaluated at u , is a continuous function of $u \in \mathbb{R}$. Therefore

$$P(\theta < \hat{\theta}_l) \rightarrow 1 - P[\Phi\{r_2(G, W)\} \leq 1 - \alpha/2], \quad \text{as } \gamma \rightarrow \infty.$$

Now consider the profile likelihood interval under \mathcal{M}_2 , with nominal coverage $1 - \alpha$. The lower endpoint of this confidence interval, denoted by $\hat{\theta}_{2l}$, is obtained by solving for $\theta < \hat{\theta}$ in

$$\Phi \left\{ r_2 \left(\frac{\hat{\theta} - \theta}{\sigma v_\theta^{1/2}}, \frac{\hat{\sigma}}{\sigma} \right) \right\} = 1 - \alpha/2.$$

Note that

$$\begin{aligned} P(\theta < \hat{\theta}_{2l}) &= P \left(\frac{\hat{\theta} - \theta}{\sigma v_\theta^{1/2}} > \frac{\hat{\theta} - \hat{\theta}_{2l}}{\sigma v_\theta^{1/2}} \right) \\ &= P \left[\Phi \left\{ r_2 \left(\frac{\hat{\theta} - \theta}{\sigma v_\theta^{1/2}}, \frac{\hat{\sigma}}{\sigma} \right) \right\} > \Phi \left\{ r_2 \left(\frac{\hat{\theta} - \hat{\theta}_{2l}}{\sigma v_\theta^{1/2}}, \frac{\hat{\sigma}}{\sigma} \right) \right\} \right] \\ &= P [\Phi\{r_2(G, W)\} > 1 - \alpha/2] \\ &= 1 - P [\Phi\{r_2(G, W)\} \leq 1 - \alpha/2] \end{aligned}$$

so $P(\theta < \hat{\theta}_l) \rightarrow P(\theta < \hat{\theta}_{2l})$, as $\gamma \rightarrow \infty$. Similarly, $P(\theta < \hat{\theta}_u) \rightarrow P(\theta < \hat{\theta}_{2u})$, as $\gamma \rightarrow \infty$ and the first part of Corollary 1 holds. The second part follows from showing that

$$\hat{\theta}_{2l} = \hat{\theta} - (n - p)^{1/2} \hat{\sigma} v_\theta^{1/2} \left\{ \exp \left(z_{1-\alpha/2}^2/n \right) - 1 \right\}^{1/2}$$

and

$$\hat{\theta}_{2u} = \hat{\theta} + (n - p)^{1/2} \hat{\sigma} v_{\theta}^{1/2} \left\{ \exp \left(z_{1-\alpha/2}^2 / n \right) - 1 \right\}^{1/2},$$

where $z_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2)$, and calculating the coverage probability of the profile likelihood confidence interval with nominal coverage $1 - \alpha$.

Proof of Theorem 2

The expected length of the MPI confidence interval, with nominal coverage $1 - \alpha$, is

$$\begin{aligned} E(\hat{\theta}_u - \hat{\theta}_l) &= \sigma v_{\theta}^{1/2} E \left\{ (\hat{\theta} - \hat{\theta}_l) / (\sigma v_{\theta}^{1/2}) - (\hat{\theta} - \hat{\theta}_u) / (\sigma v_{\theta}^{1/2}) \right\} \\ &= \sigma v_{\theta}^{1/2} E \left[\delta_{1-\alpha/2} \{ \hat{\tau} / (\sigma v_{\tau}^{1/2}), \hat{\sigma} / \sigma \} - \delta_{\alpha/2} \{ \hat{\tau} / (\sigma v_{\tau}^{1/2}), \hat{\sigma} / \sigma \} \right] \\ &= \sigma v_{\theta}^{1/2} \int_0^{\infty} \int_{-\infty}^{\infty} \{ \delta_{1-\alpha/2}(x, y) - \delta_{\alpha/2}(x, y) \} \phi(x - \gamma) f_{n-p}(y) dx dy. \end{aligned}$$

Proof of Theorem 3

Henceforth, we make the dependence of $\delta_u(x, y)$ on ρ explicit by using the notation $\delta_u(x, y, \rho)$ in place of $\delta_u(x, y)$. We first prove the following lemma.

Lemma 1. $\delta_{1-\alpha/2}(-x, y, \rho) = -\delta_{\alpha/2}(x, y, \rho)$.

Proof. From the definitions (1)–(3), we have $r_1(\delta, -x, y) = -r_1(-\delta, x, y)$, $r_2(\delta, y) = -r_2(-\delta, y)$ and $w_1(x, y) = w_1(-x, y)$. For any fixed $y > 0$, $\delta_{1-\alpha/2}(-x, y, \rho)$ is the solution in δ of

$$\begin{aligned} 1 - \alpha/2 &= w_1(-x, y) \Phi\{r_1(\delta, -x, y)\} + \{1 - w_1(-x, y)\} \Phi\{r_2(\delta, y)\} \\ &= w_1(x, y) [1 - \Phi\{r_1(-\delta, x, y)\}] + \{1 - w_1(x, y)\} [1 - \Phi\{r_2(-\delta, y)\}] \\ &= 1 - [w_1(x, y) \Phi\{r_1(-\delta, x, y)\} + \{1 - w_1(x, y)\} \Phi\{r_2(-\delta, y)\}], \end{aligned}$$

where the second line follows the fact that $\Phi(-x) = 1 - \Phi(x)$, and hence of

$$\alpha/2 = w_1(x, y) \Phi\{r_1(-\delta, x, y)\} + \{1 - w_1(x, y)\} \Phi\{r_2(-\delta, y)\}.$$

It is therefore also *minus* the solution in δ of

$$\alpha/2 = w_1(x, y) \Phi\{r_1(\delta, x, y)\} + \{1 - w_1(x, y)\} \Phi\{r_2(\delta, y)\}.$$

Since $\delta_{\alpha/2}(x, y, \rho)$ is the solution of this equation, we must have $\delta_{1-\alpha/2}(-x, y, \rho) = -\delta_{\alpha/2}(x, y, \rho)$.

Proof: *For the coverage.*

Let

$$\begin{aligned} C(\gamma, \rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ \frac{\delta_{1-\alpha/2}(x, y, \rho) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\ &\quad \left. - \Phi \left\{ \frac{\delta_{\alpha/2}(x, y, \rho) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(x - \gamma) f_{n-p}(y) dx dy. \end{aligned}$$

For each fixed ρ ,

$$\begin{aligned} C(-\gamma, \rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ \frac{\delta_{1-\alpha/2}(x, y, \rho) - \rho(x + \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\ &\quad \left. - \Phi \left\{ \frac{\delta_{\alpha/2}(x, y, \rho) - \rho(x + \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(x + \gamma) f_{n-p}(y) dx dy \\ &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ \frac{\delta_{1-\alpha/2}(-z, y, \rho) - \rho(-z + \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\ &\quad \left. - \Phi \left\{ \frac{\delta_{\alpha/2}(-z, y, \rho) - \rho(-z + \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(-z + \gamma) f_{n-p}(y) dz dy \\ &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ -\frac{\delta_{\alpha/2}(z, y, \rho) - \rho(z - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\ &\quad \left. - \Phi \left\{ -\frac{\delta_{1-\alpha/2}(z, y, \rho) - \rho(z - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(z - \gamma) f_{n-p}(y) dz dy \\ &= C(\gamma, \rho). \end{aligned}$$

The second line follows by changing the variable to $z = -x$, the third follows from the Lemma and the fact that the standard normal density is an even function, and the fourth follows from the fact that $\Phi(-x) = 1 - \Phi(x)$.

Using (4), it is straightforward to show that $\delta_u(x, y, -\rho) = \delta_u(-x, y, \rho)$. Thus,

for each fixed γ ,

$$\begin{aligned}
C(\gamma, -\rho) &= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ \frac{\delta_{1-\alpha/2}(-x, y, \rho) + \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\
&\quad \left. - \Phi \left\{ \frac{\delta_{\alpha/2}(-x, y, \rho) + \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(x - \gamma) f_{n-p}(y) dx dy \\
&= \int_0^\infty \int_{-\infty}^\infty \left[\Phi \left\{ -\frac{\delta_{\alpha/2}(x, y, \rho) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right. \\
&\quad \left. - \Phi \left\{ -\frac{\delta_{1-\alpha/2}(x, y, \rho) - \rho(x - \gamma)}{(1 - \rho^2)^{1/2}} \right\} \right] \phi(x - \gamma) f_{n-p}(y) dx dy \\
&= C(\gamma, \rho),
\end{aligned}$$

where the second line follows from the Lemma and the third from the fact that $\Phi(-x) = 1 - \Phi(x)$.

Proof: *For the expected length.*

The expected length and the scaled expected length are proportional to

$$L(\gamma, \rho) = \int_0^\infty \int_{-\infty}^\infty \{ \delta_{1-\alpha/2}(x, y, \rho) - \delta_{\alpha/2}(x, y, \rho) \} \phi(x - \gamma) f_{n-p}(y) dx dy.$$

For each fixed ρ ,

$$\begin{aligned}
L(-\gamma, \rho) &= \int_0^\infty \int_{-\infty}^\infty \{ \delta_{1-\alpha/2}(x, y, \rho) - \delta_{\alpha/2}(x, y, \rho) \} \phi(x + \gamma) f_{n-p}(y) dx dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{ \delta_{1-\alpha/2}(-z, y, \rho) - \delta_{\alpha/2}(-z, y, \rho) \} \phi(-z + \gamma) f_{n-p}(y) dz dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{ -\delta_{\alpha/2}(x, y, \rho) + \delta_{1-\alpha/2}(z, y, \rho) \} \phi(z - \gamma) f_{n-p}(y) dz dy \\
&= L(\gamma, \rho).
\end{aligned}$$

The second line follows by changing the variable to $z = -x$ and the third follows from the Lemma and the fact that the standard normal density is an even function.

It follows from $\delta_u(x, y, -\rho) = \delta_u(-x, y, \rho)$ that

$$\begin{aligned}
L(\gamma, -\rho) &= \int_0^\infty \int_{-\infty}^\infty \{ \delta_{1-\alpha/2}(-x, y, \rho) - \delta_{\alpha/2}(-x, y, \rho) \} \phi(x - \gamma) f_{n-p}(y) dx dy \\
&= \int_0^\infty \int_{-\infty}^\infty \{ -\delta_{\alpha/2}(x, y, \rho) + \delta_{1-\alpha/2}(x, y, \rho) \} \phi(x - \gamma) f_{n-p}(y) dx dy \\
&= L(\gamma, \rho),
\end{aligned}$$

where the second line follows from the Lemma.

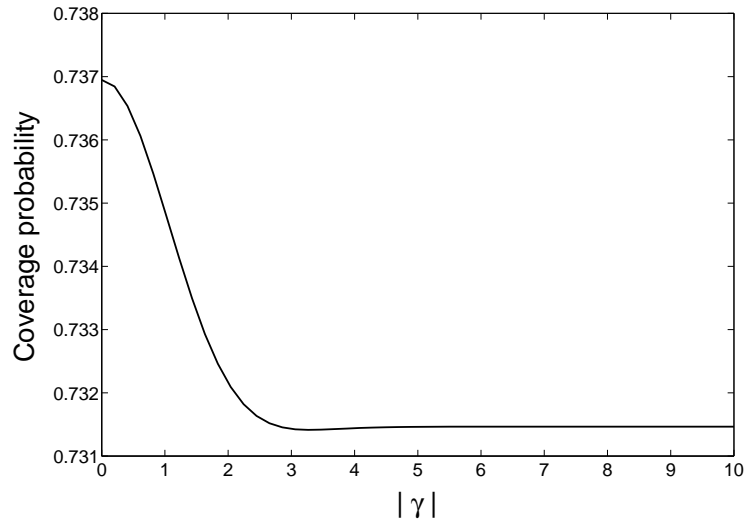


Figure 1: Plot of the coverage probability for the MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

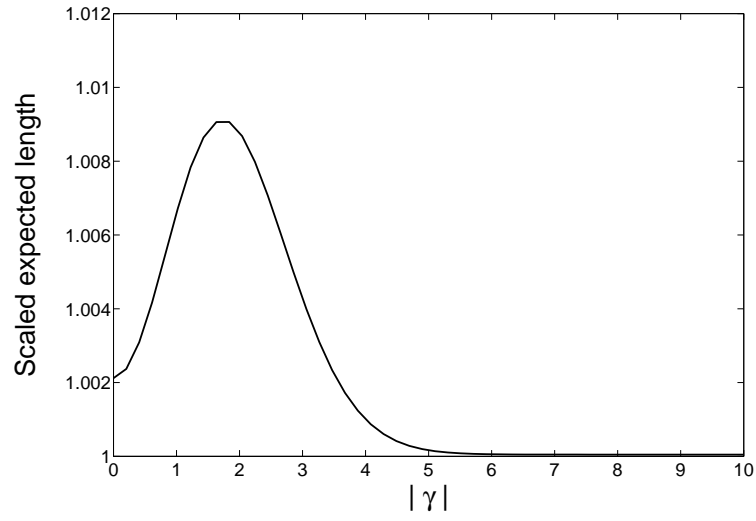


Figure 2: Plot of the scaled expected length for the MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

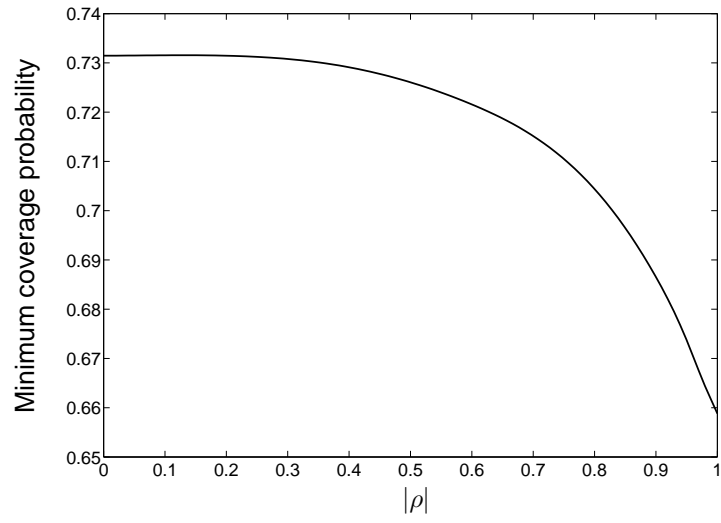


Figure 3: Plot of the minimum coverage against $|\rho|$ for the MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the squared seedability equal to zero.

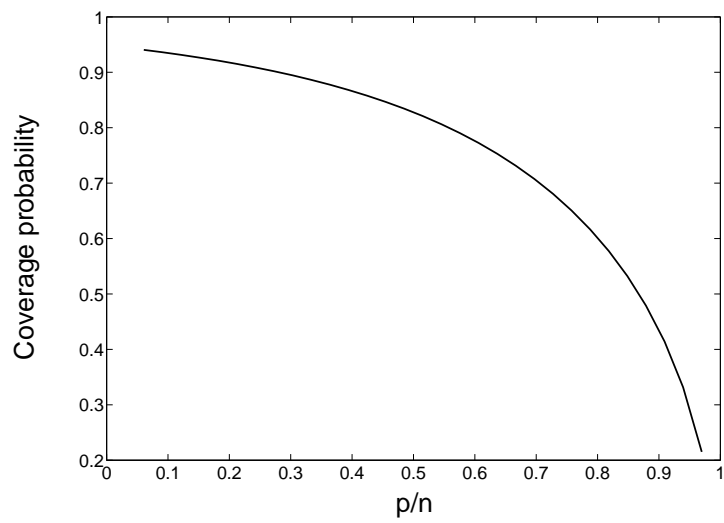


Figure 4: Plot of the coverage probability of the profile likelihood confidence interval under \mathcal{M}_2 , with nominal coverage 0.95, against p/n when $n = 33$.

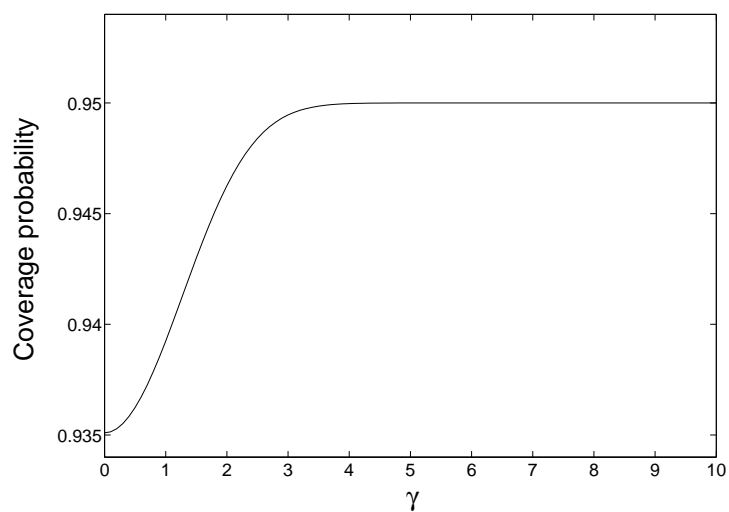


Figure 5: Plot of the coverage probability for the post-model-selection confidence interval, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the possible models are the full model and the submodel defined by setting the coefficient of the squared seedability equal to zero. The model selected is the model with smaller AIC.

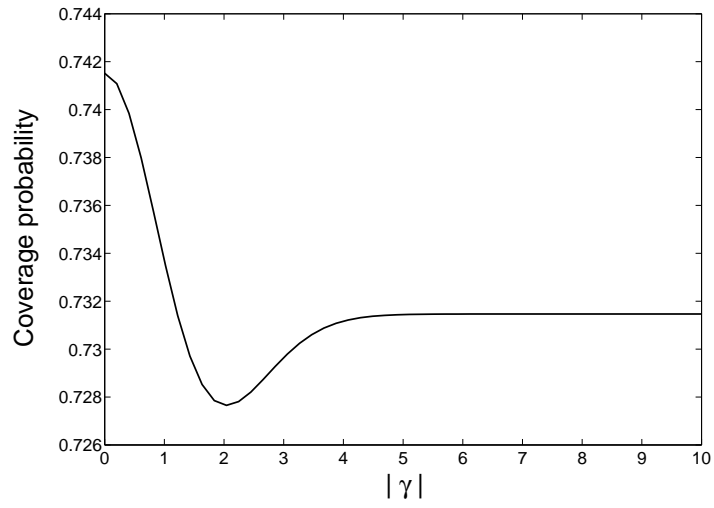


Figure 6: Plot of the coverage probability for the MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the seedability-earliness interaction equal to zero.

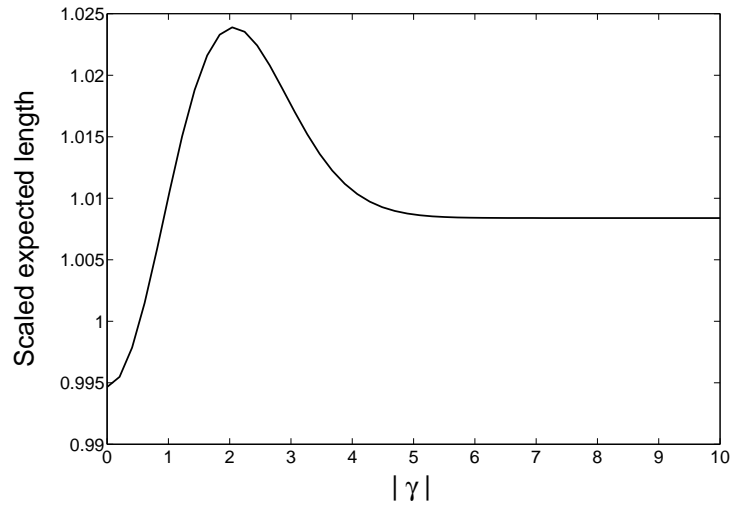


Figure 7: Plot of the scaled expected length for the MPI, with nominal coverage 0.95, for the seeding effect in the cloud seeding example when the submodel is defined by setting the coefficient of the seedability-earliness interaction equal to zero.